

Diesel Engine Valve Clearance Fault Diagnosis Based on Features Extraction Techniques and FastICA-SVM

Ya-Bing Jing^{1,3} · Chang-Wen Liu¹ · Feng-Rong Bi¹ · Xiao-Yang Bi² · Xia Wang¹ · Kang Shao¹

Received: 9 March 2016/Revised: 9 April 2017/Accepted: 20 April 2017/Published online: 15 May 2017
© Chinese Mechanical Engineering Society and Springer-Verlag Berlin Heidelberg 2017

Abstract Numerous vibration-based techniques are rarely used in diesel engines fault diagnosis in a direct way, due to the surface vibration signals of diesel engines with the complex non-stationary and nonlinear time-varying features. To investigate the fault diagnosis of diesel engines, fractal correlation dimension, wavelet energy and entropy as features reflecting the diesel engine fault fractal and energy characteristics are extracted from the decomposed signals through analyzing vibration acceleration signals derived from the cylinder head in seven different states of valve train. An intelligent fault detector FastICA-SVM is applied for diesel engine fault diagnosis and classification. The results demonstrate that FastICA-SVM achieves higher classification accuracy and makes better generalization performance in small samples recognition. Besides, the fractal correlation dimension and wavelet energy and entropy as the special features of diesel engine vibration signal are considered as input vectors of classifier FastICA-SVM and could produce the excellent classification results. The proposed methodology improves the accuracy of feature extraction and the fault diagnosis of diesel engines.

Keywords Feature extraction · Diesel engine valve train · FastICA · PCA · Support vector machine

1 Introduction

To diagnose faults of machinery under rotating, researchers around the world have attempted to make good use of vibration-based techniques and their efforts make researches. There appear many methods which have been proved to be potential in recognizing the faults of machines such as statistical analysis and time domain analysis. However, when it comes to diesel engines, it will be another case—there is very few situation using vibration-based technique to diagnose faults of diesel engines. On one hand, the vibration signals from diesel engines are momentary and unsteady. And if it is not performing, there will be no useful features extraction to diagnose. On the other, under most circumstances, diesel engines consist of reaction of several exciting forces. And the movement of diesel engine mechanism is related to the movement of many connection units such as piston ring set and valve train system. Valves, as the foremost moving unit in diesel engine mechanism, has imposed direct influence on the main noise source of diesel engine. Due to their complicated composition structure and time-to-time effects on valve seats, the fault ratio of this part occupies 15% among the faults of whole diesel engine [1].

Besides, the operation of a diesel engine will generate multiple pulse forces between valve and its sets. The vibration of cylinder head will be affected to a large degree. In case of faults on valve train, the vibration signals coming from cylinder head should consist of the fault information of it.

Supported by National Science and Technology Support Program of China (Grant No. 2015BAF07B04).

✉ Feng-Rong Bi
fr_bi@tju.edu.cn

¹ State Key Laboratory of Engines, Tianjin University, Tianjin 300072, China

² School of Mechanical Engineering, Tianjin University, Tianjin 300072, China

³ Internal Combustion Engine Research Institute, Tianjin University, Tianjin 300072, China

Over the past several years, numerous researchers have recognized the above-mentioned features and they have also conducted in-depth study on how to use vibration signals obtained from cylinder head to diagnose the fault of valve train [2–5]. Several unsteady signal analysis methods such as wavelet analysis were resorted to analyzing the unsteady vibration signals from cylinder head. Also, the fault of valve train was confirmed after analyzing more extracted features from analysis report. Such kind of diagnosis process has been widely accepted and applied in the industry in recent decades. However, it is still necessary to explore further method to extract features in a more precise and effective way. There is another difficulty that should not be ignored, that is, how to make it easier to determine how many and which features should be extracted considering the unsteady and non-linear property of vibration signals from diesel engine. Nonetheless, the problem now facing the industry is that during the deeper feature extraction process, some essential time–frequency factors possibly get lost. As a result, the diagnose result for most cases is not effective and it will be great useful to find a more impressive way.

In order to process the special vibration signal produced by diesel engine, a series of leading techniques concerning feature extraction have been proposed by researchers across the world. Time–frequency analysis, proposed by Geng, et al. [4, 5], was among many other techniques applied in extracting features. In addition, there is also another technique depending on recognizing source signals among other vibration signals gathered. Mean field independent component analysis (ICA), another application to diesel engine status monitoring, also appeared in the report of Li, et al. [6], extracted fault feature of rolling bearing based on an improved cyclical spectrum density method. Zhang, et al. [7], demonstrated the potential for applying the method to machinery fault diagnosis and the method was implemented to rolling bearing experimental data. The results obtained by using the method were consistent with the theoretical interpretation, which proved that this algorithm had imposed significant engineering significance in revealing the correlation between the faults and relevant frequency features. Chen, et al. [8], proposed an improved CICA algorithm named constrained independent component analysis based on the energy method (E-CICA) in order to realize single channel compound fault diagnosis of bearings and improve the diagnosis accuracy. Gao, et al. [9], combined the concepts of time–frequency distribution (TFD) with non-negative matrix factorization (NMF), and proposed a novel TFD matrix factorization method to enhance representation and identification of bearing fault.

However, this approach requires reference data which cannot be obtained due to the complicated external disturbance and the internal exciting forces. In 2010, two

researchers Klinchaeam and Nivesrangsan [10] analyzed one small size petrol engine with 4 stoke and 1 cylinder using vibration signal on the basis of time domain, crank angle domain, and signal energy. Yet, it is also demanded to use more powerful techniques to improve the accuracy of an engine state monitoring system as well as the accuracy of the results.

If the faults occur to the engine, the vibration signals of engine normally take on the modulation symptom under the influence of regular impulse. Adequate fault information will be included in the amplitude envelope, phase and instantaneous frequency(IF) of modulated signals. Therefore, it can be concluded that for fault extraction, fault diagnosis and fault position recognition, demodulation will serve as the precondition [11, 12]. At present, the mainly applied demodulation theory contains Hilbert transform demodulation, energy operator demodulation and so on [13–15].

However the unsteady negative IF appeared here and there, which proved that when carrying out Hilbert transform process, EMD cubic splines will lead to the loss of amplitude and frequency information. That is why local mean decomposition was presented and applied in the current days [16]. Smoothed local means are adopted by LMD to find out a more dependable IF from the vibration signals, with no factor of Hilbert transform. Wang, et al. [17, 18], had demonstrated that how LMD facilitated enhanced analysis in contrast to EMD in rub-impact fault diagnosis. In 2005, Smith [16] introduced a new suitable time-frequency analysis method named local mean decomposition(LMD). In this method, LMD, this complicated multiple component modulation signal would be disassembled into a limited set of mono-components. These components would be in linear combination and were titled as product function(PF). One envelope signal and one a frequency demodulation signal with the same amplitude would lead to the formation of one PF. The former was the momentary amplitude of PF, including amplitude modulation information. The open stage of the later with same frequency was called IF or PF, which included frequency modulation information. The LMD method was not subject to the constraint of Bedrosian theorem and Nuttall theorem due to it did not use HT at all [19].

Therefore it avoids negative frequency. However, the LMD is affected by the signal noise and many attempts have been made to extract information from noisy vibration signal. However, it was found that the current methods in the research area have limitations on the existence of noise. For example, wavelet de-noising method is completely dependent on the selection of wavelet functions. Although this method can improve a single type of noise and strong energy data when noise energy is weak and the number is

large, the capacity of noise identification and de-noising is often very poor [20]. As an effective method, Mathematical morphology is adopted to analyze geometrical structure in a quantitative way.

In this paper, the main applications of mathematical morphology in diesel engine system signal processing are concentrating on signal de-noising [21–23]. For the measurement and extraction of specific signal features, these methods make use of certain types of structure factors, with no influence from time-frequency domain; and thus have very powerful inspection and recognition capability. The mathematical morphology filter, replacing the traditional low-pass filter and mean filter, is adopted and more precise magnitude and faster response speed are obtained without nonlinear character changed.

In addition, fractal correlation dimension has become one of the most influential perspectives to recognize and predict complicated non-linear vibration behaviors. It is also widely used in engineering industry as an effective instrument to analyze signals. The diagnosis of rolling element-bearing defects has some interrelation, explained by Logan and Mathew [24], who proved that three main rolling factor bearing faults could be classified by correlation dimension, namely, outer race fault, inner race fault and roller fault. Besides the application of correlation dimension to gearbox status monitoring is also reported by Jiang, et al. [25], who illustrated that the operation status of a gearbox with wearing slit or broken teeth in contrast to normal status could be classified clearly by correlation dimension. Correlation dimension as reported by Wang, et al. [26], could not only present some inner date of a basic dynamic system, but also recognize different fault types.

The wavelet transform is considered as a popular technique to extract signal features. Wavelet functions consist of a series of basic functions describing one signal in both frequency and time perspective [27]. To make wavelet transform perform better, one new concept, wavelet energy and wavelet entropy has been brought into feature extraction. The method is widely used in many fields, and has also achieved considerable success [28–32]. On the basis of wavelet entropy, one famous method of fault diagnosis was presented by Yu, et al. [33], in which, the definition of instantaneous wavelet energy entropy (IWEE) and instantaneous wavelet singular entropy (IWSE) were derived from the wavelet entropy theory before. Later the test of this method via a real micro gas turbine engine proved the efficiency of this presented method in analyzing sensor faults.

From above, fractal correlation dimension, wavelet energy and wavelet entropy can be used to classify different faults as vibration signals features respectively, but the accuracy and efficiency maybe different. In this paper, the author attempts to apply a better combined method to

reflect the complex non-stationary and nonlinear time-varying features of the diesel engines surface vibration signals.

Independent component analysis, which is originally aimed at signal processing application has been generalized for features extraction. Principal component analysis (PCA) turns linearly the original inputs into new features with no correlation between each other while ICA turns linearly original inputs into features with mutually statistically independence with each other.

In addition, there appear several intelligent classification algorithmic rules (i.e. artificial neural network (ANNs) and support vector machines (SVMs)). These rules have been applied to the intelligent fault diagnosis of machinery and also achieved great success [34–36].

The neural network classification model has poor generalization ability due to it only conducts the experience risk minimization principle. Support vector machine (SVM) is considered as an effective machine learning method which has been used to sort out problems could lead to better generalization performance, exceeding other traditional methods. In this work, FastICA-SVM [37] is used for diesel engine valve clearance faults diagnosis. The basic concept of the methodology proposed is to apply FastICA for reduction of dimensionality and then the main influence components are input into the classifier SVM as the features to complete the diesel engine valve train faults classification.

In this paper, the author attempts to analyze the classification performance of a wide range of features. It is also discussed the combination of features on different fault data sets. To pursue the improvement of SVM classification performance, this paper brings in 5 statistical features (including mean, standard deviation, kurtosis, skewness, RMS) and 21 special features (including 5 fractal correlation dimension, 16 wavelet energy and entropy) for SVM training and explores classification performance of fractal dimensions cooperated with time-domain statistical features.

In addition, the performances of PCA-SVM and FastICA-SVM for distinguishing between different features are compared, and the effectiveness of proposed methodology is also investigated by comparison results.

In this paper, several advanced or improved methods are applied to extract features of diesel engine vibrate signal. Then identify different signal and fault types are used to obtain more accurate fault diagnosis results. As a result, the main purpose of this paper is to improve the accuracy of feature extraction as well as the fault diagnosis.

This paper is presented as below. In Sect. 2, it briefly introduces the methods principles for extracting features and classification tools. Section 3 mainly shows the experiments and data. In Sect. 4, it analyzes classification

performance of a wide range of normal and faults features extracted from various fault data. In Sect. 5, it describes and discusses the classification performance of proposed method. Section 6 is the conclusion part of whole paper.

2 Principles of Methods

2.1 Mathematical Morphology Filter

The primary concepts of mathematical morphology consist of two points: one is to find out the relationship among parts of signals and the other is to use a “probe” structural element to extract the primary signal features in circumstance of constant motion.

Basic mathematical morphological operations are defined as erosion and dilation. The opening and closing operations are derived from different combinations of erosion and dilation. The following are the definition of these operations:

Let input sequence $f(n)$ and $g(n)$ is discrete function defined in $F = (0, 1, \dots, N-1)$ and $G = (0, 1, \dots, M-1)$, respectively, $N \geq M$. The gray-scale dilation and erosion of $f(n)$ defined by $g(n)$ is as follows:

$$(f \oplus g)(n) = \max\{f(n-x) + g(x) : x \in G\}, \quad n = 0, 1, \dots, N, \tag{1}$$

$$(f \ominus g)(n) = \min\{f(n+x) - g(x) : x \in G\}, \quad n = 0, 1, \dots, N. \tag{2}$$

Set $g(n)$, in mathematical morphology, is titled as structural element, or it can be called as “probe”. The graphical features in each region can be recognized by structural element locally. Signal feature can be extracted by moving structural element constantly and we can use this information for analysis and description.

The hole of graphics can be filled by dilation operation, while the edge of a small protruding part of the graphics erosion operation can be eliminated, which smooth signal in some extent. Dilation operation is not in contrast to erosion operation. Through combining erosion and dilation, other morphological operations can be formed: opening and closing defined as follows:

$$(f \circ g)(n) = (f \ominus g \oplus g)(n), \tag{3}$$

$$(f \bullet g)(n) = (f \oplus g \ominus g)(n). \tag{4}$$

Opening operation is capable of filtering peak noise above signal and eliminates burr while closing operation is talented at smoothing trough noise behind signal classifying litter groove structure.

To explore suitable filtering order, these two mentioned operations are usually used at the same time only in

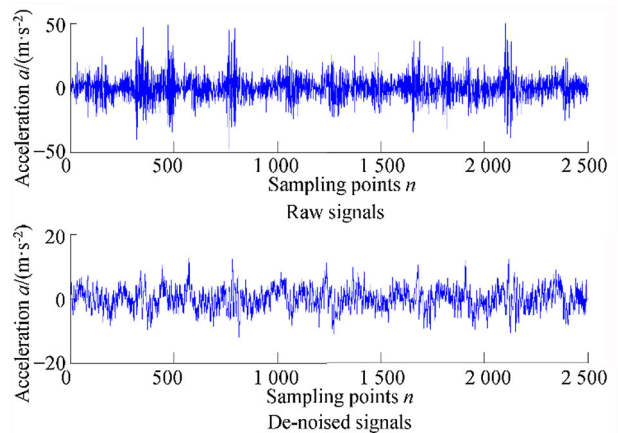


Fig. 1 Effect of mathematical morphology filter

different sequences. There is no requirement to consider the spectrum characteristics when using this method to de-noise the strong noise signals for this algorithm of mathematical morphology depends only on the local characteristics of signals to be processed.

The following Fig. 1 shows the influence of signal morphological filtering with white noise added.

2.2 Local Mean Decomposition (LMD)

In recent years, LMD was applied to separate modulated signals into small sets of PF . Each PF comes from the combination of one amplitude envelope signal and one frequency-modulated (FM) signal. LMD works as below: separating one appointed signal into frequency modulated signals and envelope components, which is defined as local magnitude functions. The below paragraphs give a brief description of real-valued LMD algorithm [38].

- (1) Based on the original signal $x(t)$, mean value, $m_{i,k}$ can be determined by calculating the mean of the successive maximum and minimum $n_{k,c}$ and $n_{k,c+1}$, where c is the index of the extrema. “ i ” and “ k ” represents the order of PF and the repetition No. in PF process. The difference between the successive extrema will decide the local magnitude, $a_{i,k}$.

$$m_{i,k,c} = \frac{n_{k,c} + n_{k,c+1}}{2}, \quad a_{i,k,c} = \frac{|n_{k,c} - n_{k,c+1}|}{2}. \tag{5}$$

- (2) Between $m_{i,k}(t)$ and $a_{i,k}(t)$, add straight lines of local mean and local magnitude values.
- (3) With moving average filter $\tilde{m}_{i,k}(t)$ and $\tilde{a}_{i,k}(t)$, remove the added local mean and local magnitude.
- (4) Deduct the removed mean signal from the original signal $x(t)$:

$$h_{i,k}(t) = x(t) - \tilde{m}_{i,k}(t). \tag{6}$$

- (5) Get $s_{i,k}(k)$, frequency modulated signal:

$$s_{i,k}(t) = \frac{h_{i,k}(t)}{\tilde{a}_{i,k}(t)}. \tag{7}$$

- (6) Confirm if $s_{i,k}(t)$ is a normalized frequency modulated signal whereas $\tilde{a}_{i,k}(t)$ is close to 1. If so, go to step (9).
- (7) If $s_{i,k}(t)$ is not a normalized frequency modulated signal, calculate $\tilde{a}_{i,k}(t) \times \tilde{a}_{i,k-1}(t)$ and return to step 1 for the same process for $s_{i,k}$.
- (8) Envelope function, $\tilde{a}_i(t)$, can be calculated by means of multiplying all $\tilde{a}_{i,k}(t)$ equals one:

$$\begin{aligned} \tilde{a}_i(t) &= \tilde{a}_{i,1}(t) \times \tilde{a}_{i,2}(t) \times \tilde{a}_{i,3}(t) \times \dots \times \tilde{a}_{i,l}(t) \\ &= \prod_{q=1}^l \tilde{a}_{i,q}(t), \end{aligned} \tag{8}$$

where l is maximum iteration number.

- (9) The envelope function, $\tilde{a}_i(t)$, is applied hereto, which multiplies the final frequency modulated signal, $s_{i,i}(t)$, to determine PF_i :

$$PF_i = \tilde{a}_i(t) \times s_{i,i}(t). \tag{9}$$

- (10) Subtract $PF_i(t)$ from $x(t)$:

$$u_i(t) = x(t) - PF_i. \tag{10}$$

- (11) Phase can be obtained from frequency modulated signal:

$$\phi_i(t) = \arccos(s_{i,i}(t)). \tag{11}$$

- (12) Instantaneous frequency (IF) will be determined by the unwrapped phase data and its differentiation:

$$w_i(t) = \frac{d\phi_i}{dt}. \tag{12}$$

First of all, there are two steps to acquire the local mean of signals. One adding mean values of successive extrema with piecewise constant interpolation and two applying moving average filter. The difference between LMD and EMD lies in how to determine the local mean function. Local magnitude function stands for IA. Frequency modulated signal results in IF directly with no necessity to adopt Hilbert transform and analytic representation. LMD uses smoothed local means and local magnitudes while EMD adopts cubic spline approach [38]. Compared with each other, smoothed local means and local magnitudes will support the natural decomposition in a better and effective way. That is why IA and IF from LMD calculation tend to be more steady and accurate.

2.3 Fractal Correlation Dimension

An engine mechanism will occupy a range of freedom degrees and the correlation dimension of nonlinear dynamic system provides the estimation of its number. That is why the correlation dimension is quite beneficial for fault diagnosis. To sort out different faults of the dynamic system, there is also other internal information available. The attractors and fractal correlation dimension value in consequence will have some kind of change when the system operation strays away from normal status. To put it in another way, system status changing will induce different fractal correlation dimension value, implying that the changing state for system. Therefore, fractal characteristics of the system can be depicted by a fractal correlation dimension value from irregular signals and it is helpful for diesel engine fault detection procedure.

To estimate the correlation dimension, it is very useful to take GP algorithm [39] into consideration. If all variables are countable, GP algorithm can be directly applied to real state space created by all variables. However, the measurements of all variables in real situations will be confronted with some difficulties. For most cases, there is only one variable scalar time series which can be measured. Based on this time series only, it is not an easy job to conclude further information of the system. Given that, the theory of time delay embedding comes into use for reconstructing the state space. This is also the first step when it comes to the analysis of a dynamical system.

The time series can be described as

$$\{X_i\}, i = 1, 2, \dots, n, \tag{13}$$

where n is the length of time series.

Speaking of n -point time series, the attractor dynamics can be rebuilt by resorting to the method of delays to create multiple state space vectors via delay coordinates:

$$\begin{aligned} x_j(m, \tau) &= (x_j, x_{j+r}, \dots, x_{j+(m-1)\tau}), \\ j &= 1, 2, \dots, n - (m - 1)\tau, \end{aligned} \tag{14}$$

where m denotes the embedding dimension, τ is the time delay parameter and $\tau = k\Delta t$.

Based on above mentioned methods, the time series with n points of data are divided into n_m groups:

$$n_m = n - (m - 1)\tau, \tag{15}$$

where n_m is the points No. or coordinate vectors No. in the fractal set.

The m -dimensional hyper sphere radius is represented by Euclidean distance:

$$r_{ij}(m, \tau) = |x_i(m, \tau) - x_j(m, \tau)| = \sqrt{\sum_{k=0}^{m-1} (x_{i+kr} - x_{j+kr})^2}. \tag{16}$$

The definition of correlation integral function $C(r)$ comes from the m -dimensional reconstructed space right after the reconstruction process. $C(r)$ presents the chance that whether there are a pair of vectors whose distance is no less than r :

$$C(r) = \frac{1}{n_m(n_m - 1)} \sum_{i=1}^{n_m} \sum_{j=1}^{n_m} H[r - r_{ij}(m, \tau)], \quad i \neq j, \tag{17}$$

$$H[r - r_{ij}(m, \tau)] = \begin{cases} 1 & r - r_{ij}(m, \tau) \geq 0, \\ 0 & r - r_{ij}(m, \tau) < 0, \end{cases} \tag{18}$$

$$r = (r_{ij,\max} - r_{ij,\min}) \frac{i + 1}{p + 1}, \quad r = 1, 2, \dots, p, \tag{19}$$

where H is the Heaviside step function.

In view of $r_{ij}(m, \tau) = r_{ji}(m, \tau)$, using the following formula the same calculation can be avoided and a half of computation saved:

$$C(r) = \frac{2}{n_m(n_m - 1)} \sum_{i=1}^{n_m} \sum_{j=i+1}^{n_m} H[r - r_{ij}(m, \tau)]. \tag{20}$$

If Eq. (17) can be rewritten as

$$C(r) = \lim_{n_m \rightarrow \infty} \frac{1}{n_m^2} \sum_{i=1}^{n_m} \sum_{j=i+1}^{n_m} H[r - r_{ij}(m, \tau)], \tag{21}$$

because r is sufficiently small and the n_m is big enough, the reconstructed phase space attractor of fractal correlation dimension can be derived as

$$D_C = \lim_{r \rightarrow 0} \frac{d[\ln C(r)]}{d(\ln r)}, \tag{22}$$

or

$$D_C = \lim_{r \rightarrow 0} \frac{\ln C(r)}{\ln r}. \tag{23}$$

Obviously, $C(r)$ is apparent that $C(r)$ and point pairs number have proportioned relationship with each other. In fractal set, they are isolated by some distances. Only when a point system is confirmed as a fractal set, can it be said that the graph of $C(r)$ in logarithmic coordinates is a linear function. The slope of it is the same as the fractal correlation dimension of system.

2.4 Wavelet Energy and Wavelet Entropy

Through using MALLAT's method [40], a certain time series $x(k)(k = 1, 2, \dots, N)$ could be decomposed as

$$\begin{aligned} x(k) &= A_1(k) + D_1(k) = A_2(k) + D_2(k) + D_1(k) \\ &= A_J(k) + \sum_{j=1}^{J-1} D_j(k), \end{aligned} \tag{24}$$

where j represents scale, $D_j(k)$ stands for the detailed components at the j th scale and $A_j(k)$ represents the approximation of the j th scale.

The frequency band of $A_j(k)$ and $D_j(k)$ could be represented by

$$\begin{cases} D_j(k) : [2^{-(j+1)}f_s, 2^{-j}f_s], \\ A_j(k) : [0, 2^{-(j+1)}f_s], \end{cases} (j = 1, 2, \dots, J, k = 1, 2, \dots, N), \tag{25}$$

where f_s is the sampling frequency.

The wavelet bases which are used to disassemble the signal are orthogonal. Due to this reason, these disassembled signals could be considered as a direct estimation of local energies at different scales [40].

Therefore, the wavelet energy of detailed components at instant k and scale j could be represented by

$$E_{j,k} = |D_j(k)|^2, \quad j = 1, 2, \dots, J, k = 1, 2, \dots, N. \tag{26}$$

The wavelet energy of approximation components at instant k and scale j is defined as below in pursuit of unification:

$$E_{J+1,k} = |A_J(k)|^2, \quad k = 1, 2, \dots, N. \tag{27}$$

As a consequence, it can be described that the wavelet energy for each scale as below:

$$E_j = \sum_{k=1}^N E_{j,k}, \quad j = 1, 2, \dots, J + 1. \tag{28}$$

The definition of total wavelet energy can be concluded as follows:

$$E_{\text{tol}} = \sum_{j=1}^{J+1} E_j. \tag{29}$$

For the j th scale, the wavelet energy ratio is treated as a normalized value:

$$P_j = \frac{E_j}{E_{\text{tol}}} \left(\sum_{j=1}^{J+1} P_j = 1 \right). \tag{30}$$

P_j which is the wavelet energy ratio vector, stands for time-scale energy distribution. It can also serve as a proper instrument to recognize and describe the features of a signal in time–frequency domain.

During the process of studying and comparing the probability distribution, Shannon entropy can be considered as one valuable standard, which offers a measure of any distribution information. Till this stage, the definition of wavelet entropy can be confirmed as below considering Shannon entropy theory and wavelet energy ratio:

$$S_{WT} = S_{WT}(P) = - \sum P_j \cdot \ln[P_j]. \tag{31}$$

In a sense, the wavelet entropy can stand for the degree of signal order or disorder. Therefore, useful information concerning basic dynamical process related with measured signals can be provided by wavelet entropy.

2.5 FastICA Based Feature Extraction Method

In the ICA algorithm, the measured variables $X \in R^d$ can be conveyed as linear combination of m unknown independent components $s = (s_1, s_2, \dots, s_m)^T \in R^m$, namely

$$X = As, \tag{32}$$

where $A \in R^{d \times m}$ is the mixing matrix. ICA is used to estimate unknown A and s only from the known X .

In this way, it is of great value to explore a de-mixing matrix W which is described as

$$\hat{s} = WX. \tag{33}$$

Therefore, vector \hat{s} after reconstruction tend to be independent. W is separated into two parts, namely, dominant part W_d and excluded part W_e . The definition of I^2 illustrating the systematic part for process variation is as below:

$$I^2 = \hat{s}_d^T \hat{s}_d, \tag{34}$$

where $\hat{s}_d = W_d X$. The definition of squared prediction error (*SPE*), which is adopted to monitor the non-systematic part of process variation as

$$SPE = e^T e, \tag{35}$$

where $e = X - \hat{X}$. I_e^2 , which equals to the process variation from the excluded part. Besides, it also makes up for the wrong selection of ICs number of the dominant part. The definition of $e = X - \hat{X}$. I_e^2 as below:

$$I_e^2 = \hat{s}_e^T \hat{s}_e. \tag{36}$$

In terms of the performance result of estimating ICA, FastICA algorithm turns out to be a greatly efficient way. The fixed-point iteration scheme, which is adopted by Fast ICA algorithm, has been proved by independent tests and it can work 10–100 times faster than traditional gradient descent methods for ICA.

In addition, FastICA algorithm is also better than others due to it plays an important role in performing projection pursuit. In this case, FastICA algorithm can offer an all-purpose data analysis method which can be applied to exploratory fashion and also estimation of independent components as well. Through making comparison between PCA and Fast ICA, the latent variables of the former are deemed to be Gaussian-distributed and the I^2 , I_e^2 and *SPE*

of the latter is separated from normality assumption. Also, PCA is inclined to process Gaussian information and Fast ICA focuses on non-Gaussian information. Based on the above mentioned point and considering the fact that many production processes present non-Gaussian characteristic, this paper takes Fast ICA as a way for feature extraction.

2.6 Supporting Vector Machine (SVM)

Vapnik [41] improved an optimal separating hyper plane in case of linearly separable, leading to the formation of statistical learning theory. Then based on statistical learning theory, SVM comes into being. SVM stands for new learning methods, during the application of which, as a classifier, original input data space is included into a high dimensional dot product space. This space is titled as feature space.

Later an optimal hyper plane is determined in the feature space to bring the generalization capacity of combined classifier to max degree. Therefore, there is some corresponding relationship between the problems of non-linear in low dimensional space with the problem of linear problem in the high dimensional space.

Based on non-linear mapping function $\varphi(x)$, and the linear function sets, the input vector x is included in the high dimensional space Z :

$$f(x, a) = (\omega \cdot \varphi(x)) + b. \tag{37}$$

Feature space, in case of high dimensional, can serve as an instrument to construct the optimal classification hyper-plane. Taking below training data set into consideration:

$$(y_1, x_1), (y_2, x_2), \dots, (y_l, x_l), \quad x \in R^n, y \in \{-1, +1\}. \tag{38}$$

The separation task is solved through the following optimal problem:

$$\begin{aligned} \min & \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^l \xi_i, \\ \text{s.t.} & \begin{cases} y_i(\omega \cdot \varphi(x_i) + b) \geq 1 - \xi_i, & i = 1, 2, \dots, l, \\ \xi_i \geq 0, & i = 1, 2, \dots, l, \end{cases} \end{aligned} \tag{39}$$

where the coefficient $C > 0$ is a penalty factor, and the coefficient ξ_i is a slack factor.

With the adoption of Lagrange multipliers, the quadratic optimization problem Eq. (39) can be solved. For this reason, it can be concluded that the hyper plane decision function is as follows:

$$f(x) = \text{sgn} \left(\sum_{i=1}^l a_i y_i (\varphi(x_i) \cdot \varphi(x_j)) + b \right). \tag{40}$$

In Eq. (40), the inner product $\varphi(x_i) \cdot \varphi(x_j)$ needs to be calculated in the feature space. In 1992, Boser, et al. [42],

proposed that calculation of inner product explicitly in the feature space was not a must item. The kernel function $K(x_i, x_j)$, based on kernel function theory, can be adopted in input space to calculate the inner produce which conforms to the Mercer condition. As a result, Eq. (40) can be conveyed as below:

$$f(x) = \text{sgn} \left(\sum_{i=1}^l a_i y_i K(x_i, x_j) + b \right). \quad (41)$$

The typical examples of kernel function are polynomial kernel, radial basis function (RBF) kernel, sigmoid kernel and linear kernel. In many practical applications [43, 44], the RBF kernel obtains the highest classification accuracy rate than other kernel functions, therefore, in the current work, we mainly consider the RBF kernel.

SVM has been originally developed to solve binary classification problems. However, practical problems often have classes more than two. To effectively extend it for multi-class classification, several methods have been proposed, such as “one-against-one”, “one-against-all”, and directed acyclic graph (DAG), where Hsu and Lin [45] gave us a comparison between these methods and pointed out that the “one-against-one” method was more suitable for practical use than other methods. In this study, we adopt “one-against-one” method to identify the different faults.

2.7 FastICA-SVM

Concerning the training of SVM, the performance will be negatively affected by the irrelevant input variables. In Ref. [46], the performance of SVM with feature extraction is better than that without. If we infer as the similar way, the priority in developing SVM fault detector is feature extraction. Figure 2 presents the structure of FastICA-SVM fault detector. In the beginning, the feature extraction based on FastICA is adopted to transform high dimension dataset into a lower one. Then, the separate components extracted are adopted to count the systematic part statistic. Due to the

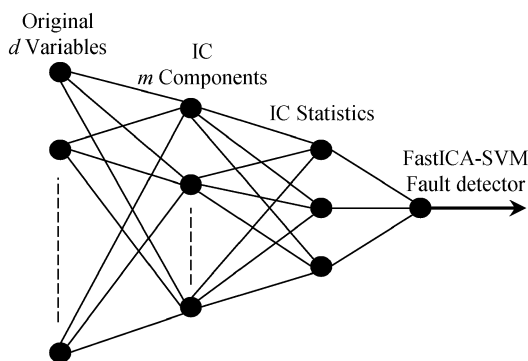


Fig. 2 Architecture of FastICA-SVM fault detector

auto-correlation of calculated statistic, timer delay and time difference of systematic statistics can be considered as input vectors for FastICA-SVM. There are mainly two phases in developing FastICA-SVM fault detector, off-line training and on-line testing. Below is the detailed procedure. Figure 3 shows the flow chart of fault diagnosis system.

2.7.1 Phase I: Off-line FastICA-SVM training

A referenced knowledge for FastICA-SVM is built in this phase and the development of normal operation condition and fault operation condition datasets are also considered.

2.7.1.1 Normal Training Dataset Development Step 1: Scale normal operation condition dataset. Obtain a normal operation condition dataset (without shifts in the process), denoted as X_{normal} . The first step focuses on centering and whitening X_{normal} , and then denote as Z_{normal} . The most cross-correlation between the observed variables are eliminated in this step.

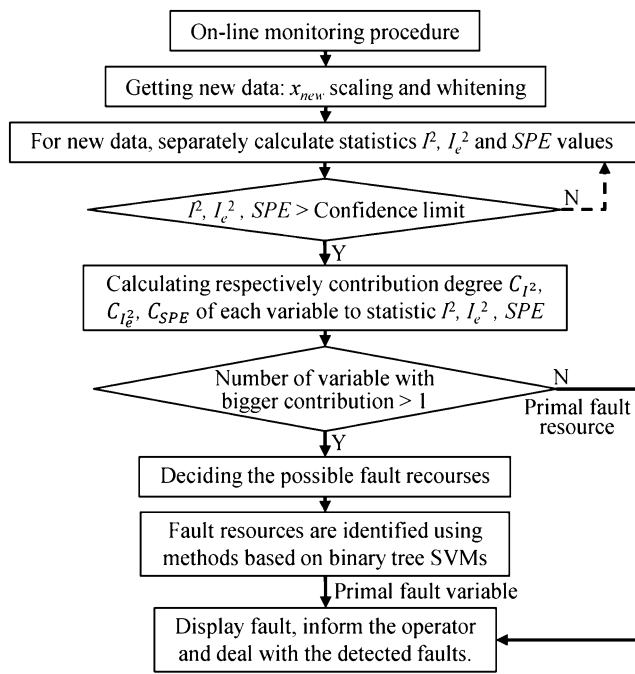
Step 2: Execute FastICA algorithm. Initially let $d = m$. Through using FastICA algorithm over Z_{normal} , and an orthogonal matrix B_{normal} can be obtained. Therefore, the reconstructed dataset is provided by $\hat{s}_{\text{normal}} = B_{\text{normal}}^T Z_{\text{normal}}$.

Step 3: Determine the order of \hat{s}_{normal} . In this step, the order of \hat{s}_{normal} is determine by using Euclidean norm (L_2) of each row(w_i) in W_{normal} : $Arg_i Max ||w_i||_2$. Therefore, a sorted de-mixing matrix can be obtained.

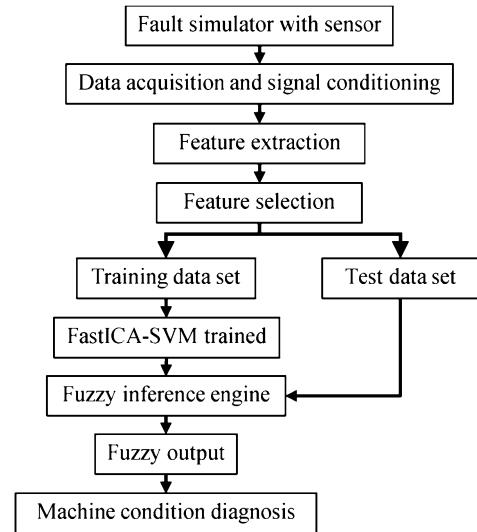
Step 4: Perform dimension reduction. Cross-validation, majority of non-Gaussianity and variance of reconstruction error and other methods can be applied to choose the separate components number. However the standard criterion to decide the number is missing. In this paper, the researcher considers the number of separate components to be the same as that of principal components.

Step 5: Calculate the systematic part statistics. The systematic part of data structure is represented by the dominant independent components. Based on the above steps, a dominant de-mixing matrix, W_d can be obtained. According to Eq. (10), $B_d = (W_d Q^{-i})^T$. Hence, the dominant independent components can be calculated by $\hat{s}_{\text{normal_d}} = B_d^T Z_{\text{normal}}$, and the systematic part statistic at sample t can be obtained, which is $I_{\text{normal}}^2(t) = \hat{s}_{\text{normal_d}}^T(t) \hat{s}_{\text{normal_d}}(t)$. The obtained I_{normal}^2 is usually auto-correlated. Hence, the time delay $I_{\text{normal}}^2(t-1)$ and time difference $I_{\text{normal}}^2(t) - I_{\text{normal}}^2(t-1)$ are additionally taken as input vectors for FastICA-SVM.

2.7.1.2 Fault Operation Condition Dataset Development Besides, fault operation condition dataset is also scaled at first, denoted as Z_{fault} , $\hat{s}_{\text{fault_d}}$ represents the



(a) Process online monitoring scheme of the proposed FastICA-SVM



(b) Fault simulator scheme of the proposed FastICA-SVM

Fig. 3 Flow chart of fault diagnosis system

dominant independent components under fault operation condition, and it can be calculated by $\hat{s}_{\text{fault_d}} = \mathbf{B}_d^T \mathbf{Z}_{\text{fault}}$. The statistic of fault operation condition systematic part at sample t is calculated from $\mathbf{I}_{\text{fault}}^2(t) = \hat{s}_{\text{fault_d}}^T(t) \hat{s}_{\text{fault_d}}(t)$. Also, $\mathbf{I}_{\text{fault}}^2(t-1)$ and $\mathbf{I}_{\text{fault}}^2(t) - \mathbf{I}_{\text{fault}}^2(t-1)$ are considered as the input vectors of FastICA-SVM.

2.7.2 Phase II: On-Line FastICA-SVM Testing

The trained FastICA-SVM model is evaluated in this phase. Once the new data is obtained, then the same scaling is applied, and the scaled dataset is denoted as \mathbf{Z}_{new} . The dominant independent components of \mathbf{Z}_{new} can be obtained from $\hat{s}_{\text{new_d}} = \mathbf{B}_d^T \mathbf{Z}_{\text{new}}$, and the statistic of systematic part at time t can be calculated by $\mathbf{I}_{\text{new}}^2(t) = \hat{s}_{\text{new_d}}^T(t) \hat{s}_{\text{new_d}}(t)$. The statistics, $\mathbf{I}_{\text{new}}^2(t)$, $\mathbf{I}_{\text{new}}^2(t-1)$ and $\mathbf{I}_{\text{new}}^2(t) - \mathbf{I}_{\text{new}}^2(t-1)$ are fed into trained FastICA-SVM for on-line process monitoring.

3 Experiment and Data Set

The vibration acceleration signals used in this paper have been acquired from the six cylinder heads in seven different states of WP7 diesel engine valve train. According to Fig. 4, the experiment test rig consists of the six-cylinder in-line diesel engine, dynamometer, LMS SCADA III multi-analyzer system (piezoelectric accelerometer sensors

and a data acquisition system with 10 kHz sampling frequency) and a dell computer. Figure 5 presents the signal-flow graph representation of engine fault diagnosis system.

The experiment comprises normal engine state and six valve clearance fault states. In normal state, all intake valve clearance is 0.3 mm and all exhaust valve clearance is 0.5 mm. In fault state, one of six cylinder was fault, the intake valve clearance is 0.4 mm and the exhaust valve clearance is 0.6 mm, as listed in Table 1 (in boldface). In the experimental work, the engine was operated in 2300 r/min with 100% load, and each sample in the fault data sets includes 2562 points.

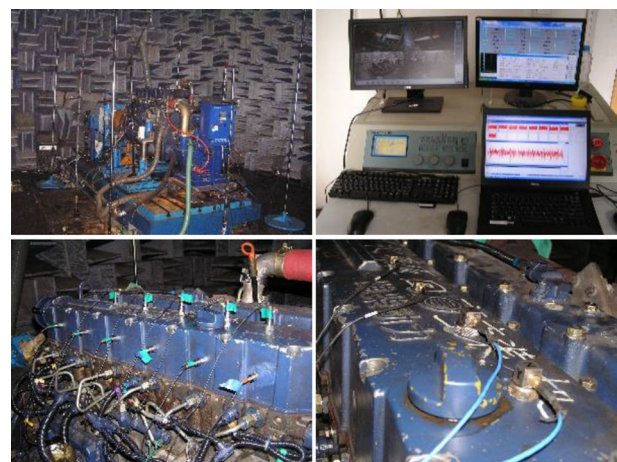


Fig. 4 Experimental test rig of diesel engine

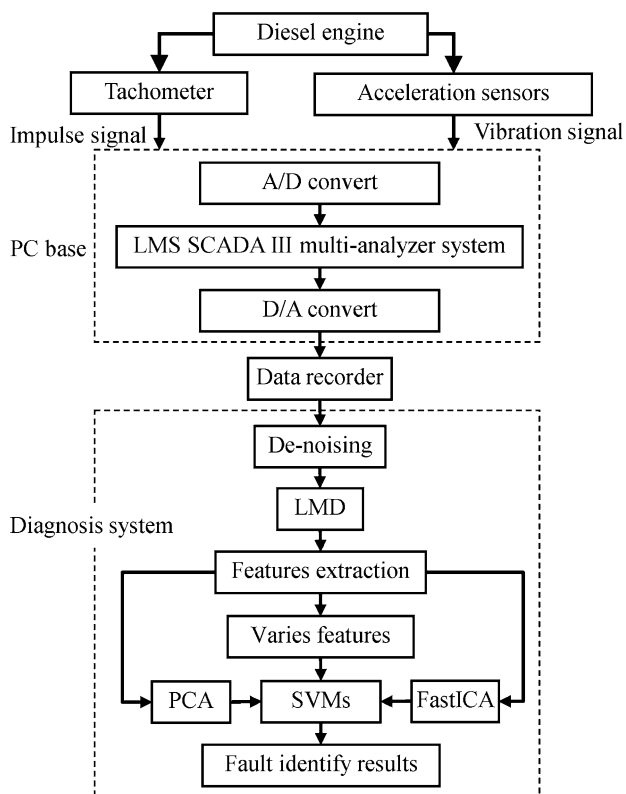


Fig. 5 Signal-flow graph representation of engine valve train faults diagnosis system

After the experimental work, the fault and normal vibration signals were recorded by 6 accelerometer sensors with a data acquisition system. And the placement of sensor specific location as below: each sensor was put on the central of each engine cylinder head. Therefore, six cylinder heads needed six sensors. Figure 5 also shows the experimental procedure and data processing of the fault diagnosis system. The acceleration sensors were used to measure the normal and fault vibration signals which were recorded with the support of a data acquisition system. The whole process as follows: the mathematical morphology filter was used to de-noise the raw signals and LMD was used to decompose the de-noised signal into a set of PFs. Since then, the statistic features, fractal correlation dimension, wavelet energy and energy of each PF was calculated as the features input into the classifiers SVM (Fast-SVM/PCA-SVM) to identify the engine fault under different operation conditions.

4 Features Extraction

4.1 Parametric Settings for Correlation Dimension Calculation

The best time delay τ and embedding dimension m are two important parameters for the phase space reconstruction.

Table 1 Seven diesel engine states

State	Valve	Valve clearance c/mm					
		cyl 1	cyl 2	cyl 3	cyl 4	cyl 5	cyl 6
State 1	Intake valve	0.4	0.3	0.3	0.3	0.3	0.3
	Exhaust valve	0.6	0.5	0.5	0.5	0.5	0.5
State 2	Intake valve	0.3	0.4	0.3	0.3	0.3	0.3
	Exhaust valve	0.5	0.6	0.5	0.5	0.5	0.5
State 3	Intake valve	0.3	0.3	0.4	0.3	0.3	0.3
	Exhaust valve	0.5	0.5	0.6	0.5	0.5	0.5
State 4	Intake valve	0.3	0.3	0.3	0.4	0.3	0.3
	Exhaust valve	0.5	0.5	0.5	0.6	0.5	0.5
State 5	Intake valve	0.3	0.3	0.3	0.3	0.4	0.3
	Exhaust valve	0.5	0.5	0.5	0.5	0.6	0.5
State 6	Intake valve	0.3	0.3	0.3	0.3	0.3	0.4
	Exhaust valve	0.5	0.5	0.5	0.5	0.5	0.6
State 7 (normal)	Intake valve	0.3	0.3	0.3	0.3	0.3	0.3
	Exhaust valve	0.5	0.5	0.5	0.5	0.5	0.5

The choice of parameters quality affects the accuracy of the correlation dimension calculation results in a direct way.

The optimum method to determine the time delay is $C-C$ method proposed by Kim et al. [47, 48], and its purpose is to select the suitable time interval as far as possible so as to keep each component in the reconstruction space independent. If the selection is too small, any of two adjacent delay coordinates $x(i)$ and $x(i+1)$ will be very numerical close, and cannot be applied as two independent coordinate systems, even leading to the information redundancy. If the selection is too big, any of two adjacent coordinates irrelevant and does not reflect the overall information system.

The purpose of choosing embedded dimension m is to make the original attractor and attractor reconstructed equivalent topology. If m chosen is too small, attractor may be fold, or even some places will appear the intersection. In some smaller areas, the points will be included on attractor different track leading to the shape of the attractor reconstructed and primitive attractor is completely different. If

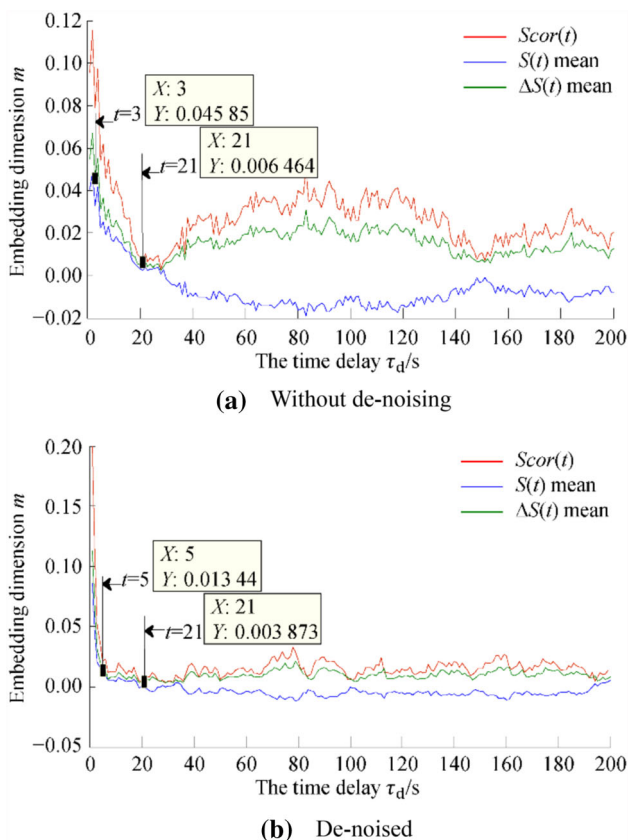


Fig. 6 Noise influence for parameters calculation

m chosen is too large, though the theory is feasible, the geometrical structure of attractor is completely opened. However, it increases the amount of calculation and the influence of the noise is also amplified. Therefore, appropriate embedding dimension is demanded to be chosen to ensure the accurate calculation and reduce the influence of calculation noise.

Due to these main parameters calculation are affected by noise seriously. A non-linear analysis procedure is applied to reduce noise level from the measured vibration signals based on the mathematical morphology filter in the first stage. The noise effect result of the fault condition vibration signal is illustrated in Fig. 6. The (a), (b) are the time delay parameter τ and time window τ_w calculated results of raw and de-noised signals measured in the first cylinder valve clearance fault condition. Figure 6 obviously indicates that the correlation dimension parameters are significantly influenced by noise level and the parameters calculated results are different.

Figure 6(a), (b) are the calculation of time delay τ_d and embedding dimension m under first cylinder valve clearance fault condition. In Fig. 6, the mean of $S(t)$ reflects the autocorrelation features of the time series, the mean of $\Delta S(t)$ measures the maximum deviation of all radius, and

the global minimum value $Scor(t)$ is the length of time sequence time window τ_w . The first minimum mean of $\Delta S(t)$ is the optimal time delay τ_d , and the global minimum value of the $Scor(t)$ is the time window length τ_w . The result shows that time delay $\tau_d = 3$ with the raw signal, while $\tau_d = 5$ with the de-noised signal, while the embedding dimensions $m = \tau_w/\tau_d + 1$ is the 8 and 6, separately.

In this work, the time delay τ_d and embedding dimension m under seven states are calculated after de-noising in Fig. 7 and the results are listed in Table 2.

4.2 Features Extracted

The total 26 features are calculated from 8 feature parameters of time domain. These parameters are respectively, Mean, Standard deviation, Root Mean Square (RMS), Skewness, and Kurtosis, 8 frequency bands Wavelet energy, 8 frequency bands Wavelet entropy, and Fractal correlation dimension values. The total of feature parameters can be shown in Table 3.

4.3 Dimensionality Reduction with PCA and FastICA

The mapping process of data from higher dimension into low dimension space can be named as feature extraction. The purpose of this process is to prevent the curse of dimensionality phenomenon. Both FastICA and PCA were adopted to decrease the feature dimensionality with 95% variation of eigenvalues. In this paper, based on the eigenvalues separately, dimensionality reduction leads to the being of 8 independents components (ICs) and 10 principal components (PCs). In addition, it can be realized that the change from data features becoming components is independent and uncorrelated. The fast independent and principal components are plotted in Fig. 8.

As for training SVM, Cao, et al. [46], had demonstrated that it was better if reducing dimensionality of SVM. Thus, in order to develop the SVM fault classifier, the dimensionality reduction based on FastICA is used for projecting the high dimension dataset into lower one. Then the extracted independent components are used to calculate the systematic part statistics, and the time delay and time difference of systematic statistics are considered as input vectors for FastICA-SVM.

Figure 9 and Table 4 to Table 6 have proved that the training and testing accuracies of SVM, PCA-SVM, and FastICA-SVM with varies features selected. The results clearly indicate that FastICA-SVM is beneficial for detecting mistakes than SVM and PCA-SVM.

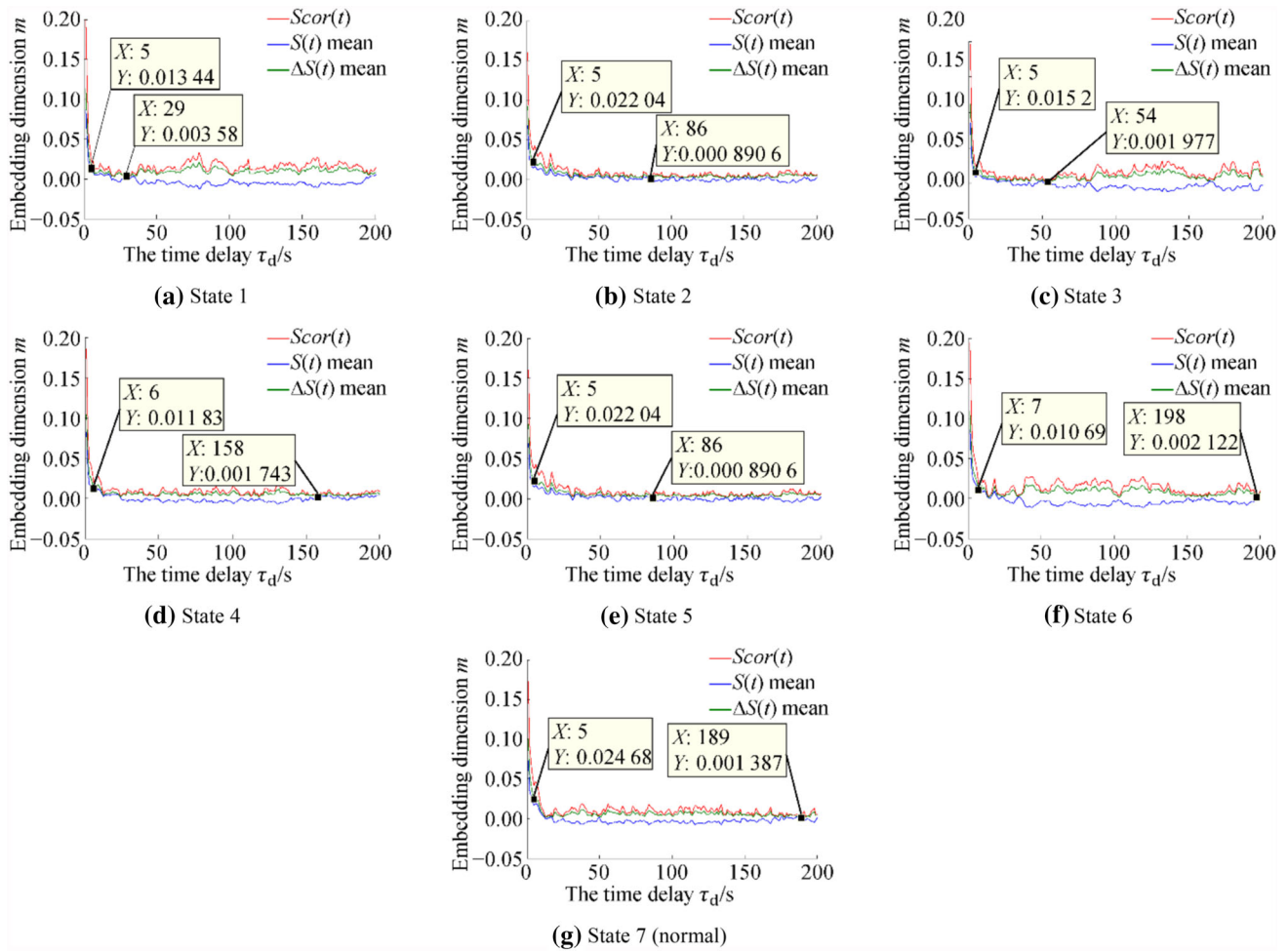


Fig. 7 Time delay τ_d and embedding dimension m under seven states

Table 2 Parameters calculation

States	Time delay τ_d/s	Embedding dimension
1	5	6
2	5	9
3	5	11
4	6	28
5	5	19
6	7	30
7	5	39

Table 3 Features extracted

Signals	Position	Feature parameters
Vibration	Vertical six cylinder heads	Mean
		Standard Deviation (SD)
		Root Mean Square (RMS)
		Skewness
		Kurtosis
		Wavelet energy
		Wavelet entropy
	Correlation dimension	

5 Results and Discussion

It shows that the result of this study in Fig. 9 and Table 4 to Table 6. The classification accuracy (%) of training and testing, the best parameters c and g for training process and the training time are listed in the above tables. The grid search method is used to select the parameters c and g because the number of grid points is not too large for only

two parameters (c and g). A possible interval of c (or g) with the grid space can be determined. Then, the cross-validation (CV) accuracy is obtained for each grid points of (c ; g). Finally, the parameters with the highest CV accuracy are returned and were used to train the whole training set. Then, the classification accuracy (%) is determined by features extracted input and the classifier type.

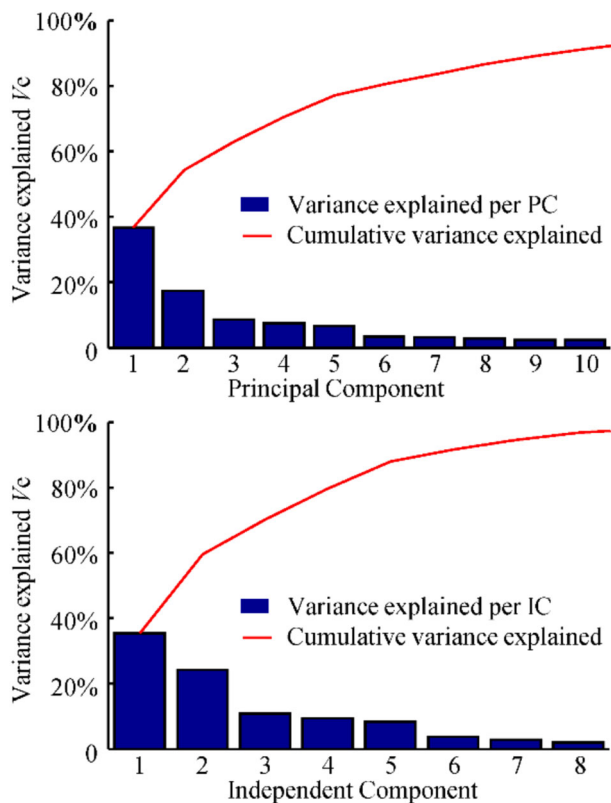


Fig. 8 Reduction of dimensions

5.1 Effect of Feature Numbers

5 statistical features (mean, SD, kurtosis, skewness, RMS) and 21 special features (5 fractal correlation dimension, 16 wavelet energy and entropy) that could explain discriminating fault conditions are as input features. Table 4 to Table 6 show the importance of their relative in classification and the efficiency of SVMs for all features. The result shows that when fractal correlation dimension combining wavelet energy and entropy as the classifier input special features, the classifier offers the best efficiency. Based on the tables, a reasonably good efficiency is obtained with fractal correlation dimension features. One can also be observed from the tables that more and more certain number of features and the efficiency falls down. It will due to consideration of some additional features that is not contributing much about classification.

There are 26 kinds of domain features extracted from decomposed vibration signals which are used as input features. It is found that the order of importance of the special features from the results, and it is considered that the features are one by one in the order from 1 feature to 26 features. The classification efficiency of SVM, PCA-SVM and FastICA-SVM are presented in Table 4 to Table 6, respectively.

The tables imply that when the number of features is 21 special features (cd + energy + entropy), the average

classification efficiency reaches its maximum value. Also, the average efficiency is good (98.286%) with all features using FastICA-SVM. The average efficiency is excellent (98.571%) When the number of features is changed to 21 special features, people generally prefer to go for 21 special features, although the reduction in efficiency is not enough 100%. People need to go for more features beyond these features mentioned in this paper if the application demands more average classification efficiency 100%. However, selecting more features more than 21 may give 100% efficiency; however, in lots of practical situations, considering the computational complexity, the preference is less compared to the all choice.

5.2 Effect of Dimensionality Reduction

Table 4 to Table 6 summarize the training and testing precision of the three ways with various features and combinations. The tables show that SVM with dimensionality reduction (PCA and FastICA) can produce higher detection rates than without dimensionality reduction. In Table 4, the classification process is performed at the primal time domain feature set without feature dimensionality reduction. The classification accuracy of this process ranges from 55.453% to 92.121%. The bad performance of this classification is due to the existence of irrelevant and useless features. Many irrelevant features make burden and tend to reduce the performance of classifier.

After that, as presented in Table 5 and Table 6, the classification rate with PCA and FastICA dimensionality reduction ranged from 80% to 98.571%. It is better than that without dimensionality reduction. Through using FastICA and PCA dimensionality reduction, the beneficial feature is extracted from primal feature sets. Furthermore, the number of support vectors (SVs) reduction due to dimensionality reduction. In this case, classification process using FastICA dimensionality reduction needs fewer numbers of SVs than PCA dimensionality reduction and original feature without dimensionality reduction. The phenomenon could be explained that FastICA finds the components not merely uncorrelated except independent. Independent components are comparatively more useful for classification rather than uncorrelated components. Because the negentropy in FastICA could take into consideration, the higher order information of the original inputs better than PCA using sample Standard deviation matrix.

5.3 Effect of Feature Selection

Table 4 to Table 6 give the classification performance of 26 time domain features and different features combinations on various data sets and make different classification

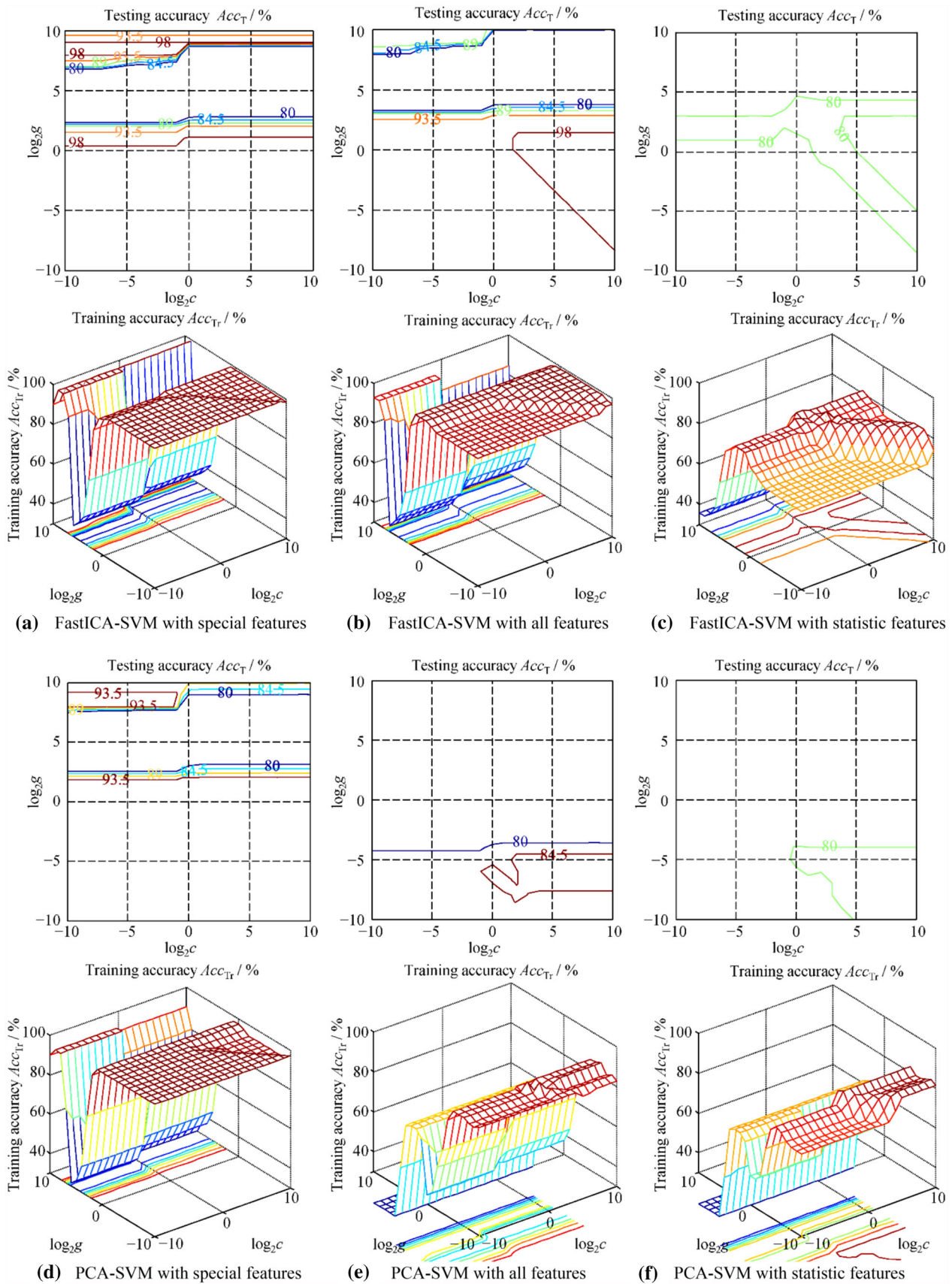


Fig. 9 Comparison of PCA-SVM and FastICA-SVM faults classifier with different features

Table 4 SVM without feature extraction classify results

Features	Best parameter (<i>c, g</i>)	Train accuracy <i>Acc_T</i> /%	Test accuracy <i>Acc_T</i> /%	Training time <i>t</i> /s
Correlation dimension	$(9.766 \times 10^{-4}, 2)$	88.416	74.473	0.312
Statistic features	(0.4, 0.5)	72.453	55.453	0.904
Wavelet energy	$(9.766 \times 10^{-4}, 3)$	85.248	64.245	0.995
Wavelet entropy	(2, 1)	78.113	72.352	0.993
Cd + Sf	(4, 0.014)	85.155	77.142	0.965
Cd + energy	(4, 1)	92.253	90.846	0.991
Cd + entropy	(16, 0.031)	84.224	75.174	0.952
Cd + energy + entropy	(8, 6)	97.432	92.121	0.985
All features	$(4, 9.766 \times 10^{-4})$	97.542	91.115	0.985

Table 5 PCA-SVM classify results

Features	Best parameter (<i>c, g</i>)	Train accuracy <i>Acc_T</i> /%	Test accuracy <i>Acc_T</i> /%	Training time <i>t</i> /s
Correlation dimension	$(9.766 \times 10^{-4}, 2)$	94.571	85.857	0.205
Statistic features	(0.5, 0.2)	81.571	80	0.894
Wavelet energy	$(9.766 \times 10^{-4}, 1)$	90.714	83.286	0.986
Wavelet entropy	(16, 1)	91.857	84.286	0.986
Cd + Sf	(4, 0.5)	90.143	82.456	0.901
Cd + energy	(4, 1)	95.246	92.857	0.986
Cd + entropy	(16, 0.031)	94.286	85.286	0.874
Cd + energy + entropy	$(9.766 \times 10^{-4}, 9.766 \times 10^{-4})$	98.571	96.714	0.969
All features	$(9.766 \times 10^{-4}, 9.766 \times 10^{-4})$	96.571	84.857	0.983

Table 6 FastICA-SVM classify results

Features	Best parameter (<i>c, g</i>)	Train accuracy <i>Acc_T</i> /%	Test accuracy <i>Acc_T</i> /%	Training time <i>t</i> /s
Correlation dimension	(4, 0.016)	97.143	87.286	0.851
Statistic features	(32, 0.016)	87.143	80	0.868
Wavelet energy	(1, 0.0166)	91.429	84.143	0.937
Wavelet entropy	(128, 0.031)	92.354	87.214	0.289
Cd + Sf	(4, 0.031)	95.714	85.143	0.747
Cd + energy	(512, 0.008)	98.547	93.571	0.188
Cd + entropy	(1, 0.063)	92.571	88.429	0.878
Cd + energy + entropy	(2, 1)	100	98.571	1
All features	(2, 0.016)	100	98.286	1

results. The classification performance increases from 80% to 98.571% with the different features and combinations. The result indicates that performance of classification accuracy is increased due to the feature selection and the classifier performance. Besides, the tables show that the classification rates of training and testing are gradually higher with the features number different and its contribution towards classification. Furthermore, the classification performance of the FastICA-SVM is highly increased through combining all of the 26 time domain features. In comparison with the 26 time domain features, because the non-linear faults information is provided to the FastICA-

SVM through fractal dimensions, wavelet energy and entropy in tandem without 5 time-domain statistical features get better classification accuracy rate on data sets. This condition should due to the bad quality of data input unable of being non-effective to represent system nonlinear characteristics well. In Table 4 to Table 6, the fractal correlation dimension is the most effective feature while the wavelet energy and entropy are second to the fractal dimension and the statistic features the worst for our classifier.

Finally, the faults classification using FastICA dimensionality reduction and all kind of features is presented in

Table 6. It presents the best performance in faults classification rather than other ways. Based on this table, it can be seen that performance of special features is 98.571% in fault classification and the feature extraction using fractal correlation dimension and dimensionality reduction with FastICA is the best method among them, due to the fractal correlation dimension can represent system characteristics effectively, while FastICA seeks not only uncorrelated, but also independent components which is more useful for classification process. Also, the fractal correlation dimension calculated after decomposing by LMD is beneficial for our fault diagnosis and the application of feature selection makes the performance of classification become more perfect.

6 Conclusions

- (1) A method which combines mathematical morphology filter, LMD, wavelet energy and entropy, and correlation dimension together presents how to extract the diesel engine faults features more accurately, which is more useful to reflect the uncertainty and nonlinear of vibration signal in time domain, and can be used to classify different faults as effective features.
- (2) The FastICA based multi-class classifier SVM is proposed to apply in faults classification process. With this method, the higher classification accuracy and better generalization performance can be achieved in small samples recognition.
- (3) Based on the proposed methods, a new fault diagnosis methodology is developed by using the features extraction, implemented via FastICA-SVM classifier. The actual example and result show that this methodology can improve the accuracy and efficiency of diesel engine fault diagnosis.

References

1. Y Xia, Z R Zhang, B L Shang, et al. Fault diagnosis for ICE based on image processing and neural networks. *Transactions of CSICE (Chinese Society for Internal Combustion Engines)*, 2001, 19(4): 356–360. (in Chinese).
2. Z Li, X C Cheng, Z B Liu. Study of diagnosis methods for diesel's valve train faults based on picture processing and neural networks. *Transactions of CSICE (Chinese Society for Internal Combustion Engines)*, 2001, 19(3): 241–244. (in Chinese).
3. H B Zheng, Z Y Li, X Z Chen, et al. Engine knock signature analysis and fault diagnosis based on time–frequency distribution. *Transactions of CSICE (Chinese Society for Internal Combustion Engines)*, 2002, 20(3): 267–272. (in Chinese).
4. Z M Geng, J Chen, J B Hull. Analysis of engine vibration and design of an applicable diagnosing approach. *International Journal of Mechanical Sciences*, 2003, 45(8): 1391–1410.
5. Z Geng, J Chen. Investigation into piston-slap-induced vibration for engine condition simulation and monitoring. *Journal of Sound and Vibration*, 2005, 282(3–5): 735–751.
6. M Li, J H Yang, X J Wang. Fault feature extraction of rolling bearing based on an improved cyclical spectrum density method. *Chinese Journal of Mechanical Engineering*, 2015, 28(6): 1240–1247.
7. K Zhang, Y Dong, A Ball. Feature selection by merging sequential bidirectional search into relevance vector machine in condition monitoring. *Chinese Journal of Mechanical Engineering*, 2015, 28(6): 1248–1253.
8. G H Chen, L F Qie, A J Zhang, et al. Improved CICA algorithm using foe single channel compound fault diagnosis of rolling bearings. *Chinese Journal of Mechanical Engineering*, 2016, 29(1): 204–210.
9. H Z Gao, L Liang, X G Chen, et al. Feature extraction and recognition for rolling element bearing fault utilizing short-Time fourier transform and non-negative matrix factorization. *Chinese Journal of Mechanical Engineering*, 2015, 28(1): 96–104.
10. S Klinchaeam, P Nivesransan. Condition monitoring of valve clearance fault on a small four strokes petrol engine using vibration signals. *Songklanakarin Journal of Science & Technology*, 2010, 32(6): 619–625.
11. Y C Choi, Y H Kim. Fault detection in a ball bearing system using minimum variance cepstrum. *Measurement Science & Technology*, 2007, 18(5): 1433–1440.
12. L L Cui, F Ding, L X Gao, et al. Research on the comprehensive demodulation of gear tooth crack early fault. *Wuhan University of science and technology journal*, 2006, 28(s2): 596–599.
13. Y Qin, S R Qin, Y F Mao. Research on iterated Hilbert transform and its application in mechanical fault diagnosis. *Mechanical Systems & Signal Processing*, 2008, 22(8): 1967–1980.
14. J S Cheng, D J Yu, Y Yang. The application of energy operator demodulation approach based on EMD in machinery fault diagnosis. *Mechanical Systems & Signal Processing*, 2007, 21(2): 668–677.
15. W Y Wang. Early detection of gear tooth cracking using the resonance demodulation technique. *Mechanical Systems & Signal Processing*, 2001, 15(5): 887–903.
16. J S Smith. The local mean decomposition and its application to EEG perception data. *Journal of the Royal Society Interface*, 2005, 2(5): 443–454.
17. Y X Wang, Z J He, Y Y Zi. A demodulation method based on improved local mean decomposition and its application in rub-impact fault diagnosis. *Measurement Science and Technology*, 2009, 20(2): 1–10.
18. Y X Wang, Z J He, Y Y Zi. A comparative study on the local mean decomposition and empirical mode decomposition and their applications to rotating machinery health diagnosis. *Journal of Vibration and Acoustics*, 2010, 132(2): 613–624.
19. Y F Dong, Y M Li, M K Xiao, et al. Analysis of earthquake ground motions using an improved Hilbert-Huang transform. *Soil Dynamics & Earthquake Engineering*, 2008, 28(1): 7–19.
20. S G Song, J T Tang, J S He. Wavelets analysis and the recognition, separation and removal of the static shift in electromagnetic soundings. *Chinese Journal of Geophysics*, 1995, 38(1): 120–128. (in Chinese).
21. P Chen, Q M Li. Design and analysis of mathematical morphology-based digital filters. *Proceedings of the CSEE*, 2005, 25(11): 60–65. (in Chinese).
22. L Ling, Z Xu. Mathematical morphology based detection and classification of dynamic power quality disturbances. *Power System Technology*, 2006, 30(5): 62–66. (in Chinese).

23. G Y Li, Y Luo, M Zhou, et al. Power quality disturbance detection and location based on mathematical morphology and grille fractal. *Proceedings of the CSEE*, 2006, 26(3): 25–30. (in Chinese).
 24. D B Logan, J Mathew. Using the correlation dimension for vibration fault diagnosis of rolling element bearing—II. Selection of experimental parameters. *Mechanical Systems and Signal Processing*, 1996, 10(3): 251–264.
 25. J D Jiang, J Chen, L S Qu. The application of correlation dimension in gearbox condition monitoring. *Journal of Sound and Vibration*, 1999, 223(4): 529–542.
 26. W Wang, J Chen, Z Wu. The application of a correlation dimension in large rotating machinery fault diagnosis. *Proceedings of the Institution of Mechanical Engineers Part C Journal of Mechanical Engineering Science*, 2000, 214(7): 921–930.
 27. I Daubechies. Ten lectures on wavelets[C]//*Society for Industrial and Applied Mathematics*, Philadelphia, PA, USA, 1992.
 28. S Blanco, A Figliola, R Q Quiroga, et al. Time–frequency analysis of electroencephalogram series. III. Wavelet packets and information cost function. *Physical Review E Statistical Physics Plasmas Fluids & Related Interdisciplinary Topics*, 1998, 57(1): 932–940.
 29. Z Y He, X Q Chen, G M Luo. Wavelet entropy measure definition and its application for transmission line fault detection and identification//*In Proceedings of 2006 International Conference on Power Systems Technology*, Chongqing, China, October 22–26, 2006: 634–639.
 30. Z Y He, S B Gao, X Q Chen, et al. Study of a new method for power system transients classification based on wavelet entropy and neural network. *International Journal of Electrical Power & Energy Systems*, 2011, 33(3): 402–410.
 31. O A Rosso, M T Martin, A Figliola, et al. Eeg analysis using wavelet-based information tools. *Journal of Neuroscience Methods*, 2006, 153(2): 163–182.
 32. W X Ren, Z S Sun. Structural damage identification by using wavelet entropy. *Engineering Structures*, 2008, 30(10): 2840–2849.
 33. B Yu, D D Liu, T H Zhang. Fault diagnosis for Micro-Gas turbine engine sensors via wavelet entropy. *Sensors*, 2011, 11(10): 9928–9941.
 34. Y Maki, K A Loparo. A neural network approach to fault detection and diagnosis in industrial process. *IEEE Transactions on Control Systems Technology*, 1997, 5(6): 529–541.
 35. B A Paya, I I Esat, M N M Badi. Artificial neural network based fault diagnostics of rotating machinery using wavelet transforms as a preprocessor. *Mechanical Systems and Signal Processing*, 1997, 11(5): 751–765.
 36. B Samant, K R Al–Balushi. Artificial neural network based fault diagnostics of rolling element bearings using time-domain features. *Mechanical Systems and Signal Processing*, 2003, 17(2): 317–328.
 37. C C Hsu, M C Chen, L S Chen. Intelligent ICA–SVM fault detector for non–Gaussian multivariate process monitoring. *Expert Systems with Applications*, 2010, 37(4): 3264–3273.
 38. C Park, D Looney, M M V Hulle, et al. The complex local mean decomposition. *Neurocomputing*, 2011, 74(6): 867–875.
 39. P Grassberger, I Procaccia. Characterization of strange attractors. *Physical Review Letters*, 1983, 50(5): 346–349.
 40. S G Mallat. A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Trans. Patt. Anal. Mach. Intell.*, 1989, 11(7): 674–693.
 41. V N Vapnik. *The nature of statistical learning theory*. New York: Springer, 1995.
 42. B E Boser, I M Guyon, V N VAPNIK. A training algorithm for optimal margin classifiers//*Proceeding of the 1992 Fifth Annual ACM Workshop*, Pittsburgh, Pennsylvania, USA, July 27–29, 1992: 144–152.
 43. Y S Zhu. *Support vector machine and its application in mechanical fault pattern recognition*. Xi’an: Jiaotong University, 2003.
 44. J Y Yang, Y Y Zhang. Application research of support vector machines in condition trend prediction of mechanical equipment. *Lecture Notes in Computer Science*, 2005, 3498: 857–864.
 45. C W Hsu, C J Lin. A comparison of methods for multi-class support vector machines. *IEEE Transactions on Neural Networks*, 2002, 13(2): 415–425.
 46. L J Cao, K S Chua, W K Chong, et al. A comparison of PCA, KPCA and ICA for dimensionality reduction in support vector machine. *Neurocomputing*, 2003, 55(1–2):321–336.
 47. D Kugiumtzis. State space reconstruction parameters in the analysis of chaotic time series—the role of the time window length. *Physica D Nonlinear Phenomena*, 1996, 95(1): 13–28.
 48. H S Kim, R Eykholt, J D SALAS. Nonlinear dynamics, delay times, and embedding windows. *Physica D Nonlinear Phenomena*, 1999, 127(1–2): 48–60.
- Ya-Bing Jing**, born in 1979, is currently a PhD candidate at *State Key Laboratory of Engine, Tianjin University, China*. He received his master degree from *Tianjin University, China*, in 2010. His research interests include engine noise and vibration signal processing, engine noise and vibration control. Tel: +86-13820278383; E-mail: robin@tju.edu.cn
- Chang-Wen Liu**, born in 1963, is currently a professor at *State Key Laboratory of Engine, Tianjin University, China*. His research interest is engine’s electronic control technology. Tel: +86-13602057489; E-mail: liuchangwen@tju.edu.cn
- Feng-Rong Bi**, born in 1965, is currently a professor at *State Key Laboratory of Engine, Tianjin University, China*. He received his PhD degree from *Tianjin University, China*, in 2003. His research interests include engine noise and vibration control, automobile dynamics, etc. Tel: +86-13802167153; E-mail: fr_bi@tju.edu.cn
- Xiao-Yang Bi**, born in 1989, is currently a PhD candidate at *School of Mechanical Engineering, Tianjin University, China*. She received her bachelor degree from *Tianjin University, China*, in 2010. Her research interests include non-linear vibration analysis and vibration control. Tel: +86-13512915589; E-mail: beckybi1001@gmail.com
- Xia Wang**, born in 1984, is currently working at *China Association of Automobile Manufacturers, China*. She received her PhD degree from *Tianjin University, China*, in 2015. Her research interests include feature extraction, signal processing and fault diagnosis. Tel: +8618612997686, E-mail: wangxia@caam.org.cn
- Kang Shao**, born in 1981, is currently working at *Tianjin Sino-German University of Applied Sciences, China*. He received his PhD degree from *Tianjin University, China*, in 2014. His research interests include engine main bearing’s hydrodynamic performance, engine crankshaft dynamic, and engine dynamics simulation. Tel: +8615822457519, E-mail: kangshao1981@tju.edu.cn